

**Description of plans for a SIPP calendar validation study:  
Study design and analysis**

Jason M. Fields, Housing and Household Economic Statistics Division, US Census Bureau  
Jeffrey C. Moore, Statistical Research Division, US Census Bureau

---

*This report is released to inform interested parties of research and to encourage discussion of work in progress.  
The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.*

---

**Abstract**

Plans for the Census Bureau's re-engineered Survey of Income and Program Participation (SIPP) program include use of event history calendar (EHC) interviewing methods, and (assuming a favorable research outcome) a 12-month, calendar-year reference period, in place of a standard questionnaire approach with a sliding 4-month reference period. This paper describes the first field test research project to compare the quality of the data obtained under the two approaches. The essential feature of the research is a small-scale field test, in early 2008, of a prototype paper EHC questionnaire, covering calendar year 2007, administered to expired 2004 panel SIPP households who will have already reported about calendar year 2007 via their final three waves of SIPP interviews. Analysis will focus on a comparison between the two interviewing methods of the reporting of key characteristics (e.g., participation in programs, jobs/businesses, and health insurance coverage), their start and stop dates, and (where relevant) income amounts. Because little is known about how EHC methods are actually put into practice in the field, the 2008 study will also employ a variety of additional evaluations -- interviewer and respondent debriefings, observations, analysis of recorded interviews, etc. -- directed toward a better understanding of the EHC interview process. Subject to available funding, the field test will be administered to two states, IL and TX. Administrative records data to validate program participation from the two survey based estimates are in the process of being obtained. Following the survey based analyses; validation evaluations will be conducted with these records.

---

## **Overview**

The US Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP) to accomplish several goals, including reducing burden on respondents, reducing program costs, improving accuracy, improving timeliness and accessibility, and improving relevance. The main objective of the SIPP has been to provide accurate and comprehensive information about the income and program participation of individuals and households in the United States. The survey's mission is to provide a nationally representative sample for evaluating: 1) annual and sub-annual income dynamics, 2) movements into and out of government transfer programs, 3) family and social context of individuals and households, and 4) interactions among these items. The survey re-engineering of SIPP pursues these objectives in the context of several goals - cost reduction and improved accuracy, relevance, timeliness, and accessibility. The SIPP collects detailed information on cash and non-cash income (including participation in government transfer programs) three times a year, and detailed data on taxes, assets, and liabilities are collected annually. A major use of the SIPP has been to evaluate the use of and eligibility for government programs and to analyze the impacts of options for modifying them.

A key component of the re-engineering process involves the proposed shift from the every-four-month data collection schedule of traditional SIPP to annual data collection in the re-engineered survey. To accomplish this shift with minimal harm to data quality, the Census Bureau proposes to employ event history calendar (EHC) methods to gather SIPP data (Fields and Callegaro, 2007). Belli (1998) provides a strong theoretical rationale for the use of EHC methods, and their likely superiority to more traditional survey instruments using a standard question-by-question approach. Most existing EHC evaluations are consistent with the hypothesis of improved data quality – by improvements in the ability of respondents to integrate memory across topic areas, and retrieve related information in a more natural autobiographical manner. The research base is somewhat limited in terms of strong quantitative evaluations of theory-based predictions. Most studies have focused on the use of comparable survey recall periods and evaluated strictly the survey method. Thus, concern lingers about the data quality implications for the topics covered in SIPP of the shift from a four-month recall period to a one-year recall period.

## **Background**

The event history calendar (EHC) is a survey methodology that has been successfully employed since the 1960's to assist interviewers in collecting detailed data with long recall periods (Belli, 1998; Belli, Shay, and Stafford, 2001; Callegaro 2007). Although never implemented as a production instrument the Census Bureau and SIPP researchers have experience with EHC instruments. In the late 1980's an EHC was field tested with SIPP in the Chicago region (Kominski, 1990). In the end this test was not implemented as a production component because there were too many concomitant changes required to integrate it into the program. In the late 1990's, EHC instruments began to be developed as electronic instruments, significantly easing some of the issues associated with retrieving and coding the data collected with this tool. The

EHC methodology helps interviewers and respondents by allowing recall of information in a more natural “autobiographical” manner. The respondent cues their memories off of landmark events they can clearly place in time during the reference period, as well as from answers to other domains in the EHC. We outline the basic information surrounding the decision to pursue an EHC in the re-engineered SIPP as well as some background on the history EHC instruments in Fields and Callegaro (2007).

As part of the development of the re-engineered SIPP we are incorporating two distinct field test activities. The first research project is a re-interview and validation experiment using a sample of cases from the SIPP 2004 panel. The purpose of this first evaluation is to consider the ability of the EHC to successfully assist respondents to recall program related information. A second evaluation will be conducted later in 2009 using an electronic EHC currently under development at the Census Bureau, and will integrate the rest of the Census Bureau survey management and processing components.

This research paper describes a project to compare the quality of data obtained under the two interviewing approaches. The essential feature of the research is a small-scale field test, in early 2008 of a prototype EHC questionnaire, covering calendar year 2007, administered to expired 2004 panel SIPP households who will already have reported about calendar year 2007 via their final three waves of SIPP interviews (see Figure 1).

Analysis will focus on a comparison between the two interviewing methods focusing on the reporting of key characteristics (e.g., participation in programs, jobs and businesses, health insurance coverage, school enrollment, and residences), their start and stop dates, and (where relevant) income amounts. Because little is known about how EHC methods are actually put into practice in the field, the 2008 study will also employ a variety of additional evaluations – interviewer and respondent debriefings, observations, analyses of recorded interviews – directed toward a better understanding of the EHC interview process. The qualitative information gained from these observations will help to refine the training and identify problems that may need to be addressed before the 2009 dress rehearsal is fielded. We also hope that these qualitative methods can be useful in understanding any differences in the quantitative data collected by the EHC from the comparison data. These evaluations will help to differentiate issues that can be corrected through training from inadequacies in the instrument or methodology.

The field test will be limited to two states, Illinois (IL) and Texas (TX). These states were chosen for ease of administration and, primarily to facilitate the use of administrative record data for a more rigorous data quality validation assessment for selected characteristics. Additionally, this test will provide these two regional offices, their management and field representatives with valuable experience with the EHC survey methodology. This experience will be invaluable as we transfer what we learn about training interviewers on this first field test EHC to the full 2009 dress rehearsal.

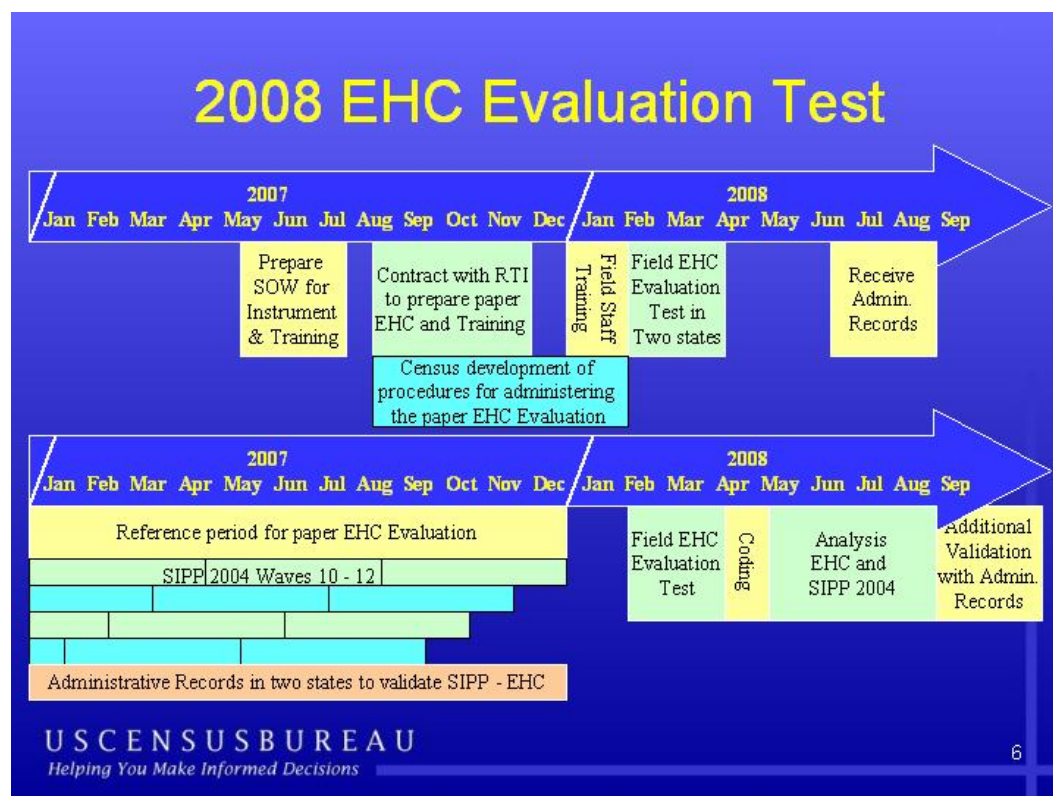
One limitation of the design is the possibility that the SIPP respondents' EHC reports will be "primed" by having just completed three waves of SIPP interviews covering the same time period. Certainly, the experience of having been SIPP respondents will predispose these respondents to being able to accurately recall they type of information we have included in this test. This "priming" would be a significant problem if there were no plan to evaluate its effect. However, this study will yield data about the effects of such priming by including all the available un-primed SIPP cases from the same states that, in a budget-cutting exercise, were dropped from the SIPP sample after wave 8, and thus who will not have previously reported about calendar year 2007. In addition to the un-primed cases, we will be able to interview new residents who moved into previously interviewed SIPP addresses.

Figure 1.

SIPP 2004 PANEL REFERENCE PERIOD MONTHS IN CALENDAR YEAR 2007 BY ROTATION GROUP								
CALENDAR MONTH	ROTATION GROUP							
	1		2		3		4	
	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month	Ref. Period	Intvw. Month
2006 October	W10	W9						
November				W9				
December			W10			W9		
<b>2007 JANUARY</b>	W11				W10		W10	W9
<b>FEBRUARY</b>		W10						
<b>MARCH</b>				W10			W11	
<b>APRIL</b>			W11			W10		
<b>MAY</b>	W12				W11		W11	W10
<b>JUNE</b>		W11						
<b>JULY</b>				W11			W12	
<b>AUGUST</b>			W12			W11		
<b>SEPTEMBER</b>	W12				W12		W12	W11
<b>OCTOBER</b>		W12						
<b>NOVEMBER</b>				W12			W12	
<b>DECEMBER</b>						W12		
2008 January								W12
** FEBRUARY 2008 – START OF NEW 2008 PANEL **								

## Research Plan Overview

This field test and evaluation is being designed to address several specific measurement and survey administration issues. The design of this study is comparative; SIPP vs. EHC (primed), and EHC (primed) vs. EHC (un-primed). While developing a plan for re-engineering SIPP and determining the revised survey content, the Census Bureau conducted numerous stakeholder briefings and meetings. During the course of these meetings one of the more common concerns that was raised was whether the proposed EHC would be able to measure program participation as well as the current SIPP design. A primary concern is that the cost savings generated by reducing the number of interviews to one per year rather than three would come at too high a cost in terms of data quality – especially in the context of program participation. The schedule of field test activities is represented in Figure 2. This paper represents the planning and status of the 2008 field test project as of September of 2007. The paper instrument and training materials are still in development and not able to be included in this paper at this time. The instrument, still in revision, will be available by the end of 2007 so that training materials and training sessions can be completed in early 2008. Subject to available funding, the field administration of this test will likely begin in March 2008. This will allow time for field activities and training for the 2008 panel of the SIPP to be started before this test is administered.



The first comparison that we will be making is simply to assess the recording of events in EHC vis-à-vis the SIPP control data (SIPP vs. EHC (primed)). Responses to the 12-month EHC will be compared with the same respondents' SIPP interview reports covering the same calendar year. Missed events in one or the other interview method are likely evidence of reduced data quality. The events being evaluated include (Key SIPP Variables Involved – Public Use Names):

1. Residential Moves (SHHADID, TFIPSST, TMETRO, RHCHANGE, EPUBHSE, EGVTRNT, EWRSECT8)
2. School Enrollment (RENRLMA, EERLM, EENLEVEL, EEDUCATE)
3. Labor Force (EBNO1, EBNO2, TBSOCC1, TBSOCC2, EENO1, EENO2, TJBOCC1, TJBOCC2, RPYPER1, RPYPER2, TPMSUM1, TPMSUM2, RMERS, ELAYOFF, ELKWRK, RWKESR1, RWKESR2, RWKESR3, RWKESR4, RWKESR5, TFUNEMP)
4. Workers Insurance Programs (ER05, ER06, ER10, ER14, EUECTYP5, EUECTYP6, T05AMT, T06AMT, T10AMT, T14AMT, EDISABL, EDISPREV)
5. Health Insurance (ECDMNTH, ECRMNTH, EHEMPY, EHIMTH, EHIOwner, EMCOCOV, RCHAMPM, RMEDCODE, RPRVHI, RPRVHI2, RCUOW58A, RCUOW58B, RCUTYP58)
6. Social Security (RCUOWN01, RCUTYP01, ER01A, ER01K, T01AMTA, T01AMTK, ECRMTH, RMEDCODE, TFSOCSEC)
7. Social Welfare Programs (RCUOWN03, RCUOWN04, RCUOWN25, RCUOWN27, RCUTYP03, RCUTYP04, RCUTYP25, RCUTYP27, TFSSI, TFTRNINC, EFSYN, EWICYN, EPATANF1, EPATANF2, EPATANF3, EPATANF4, EPATANF5, EPATANF6, ER03A, ER03K, ER04, T03AMTA, T03AMTK, T04AMT)
8. Asset Ownership (EAST2D, EAST1B, EAST2A, EGVJT, ECDJT, ECKJT, EMDJT, EBDJT, ESVJT, EAST2C, EAST3E, EMRTJNT, EMRTOWN, EAST3A, EAST3C, EAST4C, ESVOAST, EAST4A, EAST4B, EGVOAST, ECDOAST, ECKOAST, EMDOAST, EBDOAST, EAST3B, EAST3D, EAST1A)

The recording of these events will be evaluated based at multiple levels of agreement. Do respondents reporting receipt of programs in SIPP; also identify receipt of the programs in the EHC? Do the changes in status occur in the same month? Do the changes align within 1-2 months? Are simultaneous changes across domains reported consistently in both instruments? Care will be taken to consider erroneous transitions generated by SIPP seam issues, and to distinguish reported data from imputed data in the SIPP comparison data.

Other data quality differences may be suggested by the quality of the distributions of spell transitions across calendar months. This phase of the analysis will compare the levels and patterns recorded in each of the three interview components: SIPP 2004, EHC (primed), and EHC (un-primed). Comparison of the data recorded from the two groups of EHC respondents will provide a way to control for biases introduced due to the re-interview design. There will be respondents in the un-primed group that will have some baseline data – allowing background patterns of program receipt to be used in the evaluation of this group's data as well. There will be additional cases from both sample groups that represent households where the previous SIPP household members left and a replacement household (new residents at the sampled address) will be interviewed. These replacement households may provide yet another way to evaluate and control for biases associated with respondent's prior experience with SIPP.

As outlined above, the data quality analyses will focus on the measurement repeat measurement of respondents' events in both the SIPP and from the EHC. We will construct categorical outcome variables for each of the domains that indicate:

OCCUR (All cases)

1. Spell in both SIPP and EHC
2. Spell in SIPP not EHC
3. Spell in EHC not SIPP
4. No spell in SIPP or EHC

TIMING (Cases with spells in both SIPP and EHC)

1. SIPP and EHC agreement on month
2. SIPP and EHC 1-3 month difference in incidence month
3. Spell in both SIPP and EHC more than 4 months difference in timing

The percent distribution by domain in these outcome variables will be evaluated to determine where differences occur and in which direction (greater or lesser reporting of events in SIPP versus the EHC).

We will focus analyses for each domain on the relative timing during the calendar year of events. This will allow us to address concerns that the reporting of events degrades with a longer recall period. As described, the EHC is a tool to aid in recall and improve consistency over topical domains. If successful the EHC will not substantially underreport events at the beginning of the year relative to the reporting of events at the middle or end of the year. To evaluate this, we will be considering the distributions of events over the thirds of the year. Due to the rotational nature of the SIPP sample these thirds will not easily overlay the waves in the SIPP, but SIPP events and distributions can be no more than 4 months from the interview, and will still provide a good comparison even though the first third of the EHC year will overlap waves and reference months in the SIPP data.

We are careful at this point, not to identify our results in terms of better or worse for most of these comparisons. Events that occur in SIPP on seams may be erroneous; certainly the timing of these events is suspect if they are concentrated on seams. SIPP allocation rates and allocated events need to be considered in the comparison – these will be removed from some of the analyses and noted specifically in the rest of them. The validation analyses will be key in determining accuracy. Although for unallocated SIPP data, we expect that the proximity of the reporting to the event should have yielded better reporting for most events.

Distributional characteristics, such as the percent with TANF, Food Stamps, Medicare, Working, Enrolled, and with Health Insurance coverage from the EHC will be compared to the same distributions from SIPP. Once administrative records are available, programs with comparison data will be compared to the distributions we can generate from administrative records in the validation component of the analysis. We will produce indices of dissimilarity, indicating how much one distribution would have to be adjusted to mirror the other.

The inter-domain consistency will be evaluated to determine the relative timing of events across topics. We expect the EHC will significantly improve the consistency across domains, and this will be analyzed by looking at the correlations between events from different topic domains in both SIPP and EHC and see which has stronger correlations.

Additional evaluation methods – respondent debriefings, interviewer debriefings and focus groups, interview observations, analysis of recorded interviews, etc. – will be directed toward a better understanding of the EHC interview process, such as how landmark dates are introduced and used, the preferred “direction” of reporting, the extent to which events in one domain are used to pinpoint transitions in another domain, etc.

The two states, Illinois and Texas, are ideal test areas for this evaluation. There are sufficient cases from SIPP 2004 in these areas and there is solid groundwork in place to put agreements together to utilize administrative records in a validation step to the analysis. Table 1 presents the current households available to be interviewed in each area (and Maryland), and identifies them as continuing (Wave 10) households or sample-cut (Wave 8) households. If on average there are two adult respondents per household, the 1,984 households will generate nearly 4,000 individual EHC records for analysis. There should be approximately 1100 households or 2,000 individuals where we can directly compare their EHC responses to their SIPP 2004 responses. The cases in Texas are a subset of all the 2004 SIPP cases in Texas. We chose to focus on respondents in metropolitan areas to maximize the program cases available for evaluation.

The validation component of the analysis, where we compare both SIPP 2004 responses and EHC responses to administrative records depends on reaching the necessary data agreements with the administrative data sources. The first component of the analysis, re-interview and comparison of SIPP and EHC data can proceed before the data agreements are finalized. Substantial groundwork has already been laid to be able to utilize administrative records for



several programs (e.g., TANF, Food Stamps, Medicare, Social Security, SSI, and possibly wage information). This validation stage of the analysis will occur after the first stage comparisons due to the added time necessary to obtain and match the necessary administrative records.

Table 1. 2008 Field Test -- Approximate Number of Available Cases

SIPP 2004 Available Cases (1)	Illinois	Texas	Maryland
	Available Households	Available Households	Available Households
Total households	936	1048	884
Wave 10 completed households	508	614	268
Wave 8 reduced households	428	434	616

---

Source: Survey of Income and Program Participation - 2008 Re-engineering field test 1.

Notes: (1) Households were selected for interview in the field test from those completing interviews through Wave 10 in Illinois and in four metropolitan areas of Texas.

(2) If a selected address interviewed for the Event History Calendar test does not include any SIPP 2004 respondents we will utilize the cases as a type of 'un-primed' replacement households.

By including direct comparisons across survey instruments, as well as an administrative-record-based validation component, this research will be able to add significantly to the literature on event history calendar survey methodology, especially with respect to validating the SIPP and EHC reporting of income transfer program receipt and amounts over a calendar year. Results from the study will also inform the decision of whether to use EHC methods in the re-engineered SIPP program currently under development at the Census Bureau.

### Next Steps

Following the 2008 paper instrument evaluation, (assuming a positive outcome) a broad dress-rehearsal evaluation of the new electronic EHC instrument being designed for the re-engineered SIPP for possible administration in September 2009. The results from the 2008 EHC evaluation will be used to refine training procedures and make necessary adjustments to the new computer assisted personal interview (CAPI) EHC being prepared for the dress rehearsal.

The planning and instrument development for the 2009 re-engineered SIPP dress rehearsal is well underway. The survey is scheduled to be administered in September – the earliest possible administration window for the dress rehearsal. It will collect information about jobs, programs, health insurance and demographics for the 2008 calendar year. The dress rehearsal will implement the lessons learned in developing field procedures for the 2008 EHC evaluation and extend field implementation to each of the Regional Offices for this national test. The 2009 dress rehearsal instrument will be evaluated in several domains including field implementation issues and data comparability vis-à-vis SIPP 2008 and administrative records. The administration of the 2009 dress rehearsal in September is not ideal, but is the earliest in 2009 that the instrument can be ready for implementation. The production implementation of an EHC

in the re-engineered SIPP would be during the early part of the calendar year to minimize the length of recall in the reporting of data for the prior calendar year. Results from both the 2008 evaluation and the 2009 dress rehearsal will be used to make final decisions regarding the design and implementation of the re-engineered SIPP for production in 2011 or 2012.

### **Questions For Advisory Committee**

-- We seek constructive criticism concerning any and all aspects of the study design and directed toward ways to strengthen it.

-- For the primary comparison of interest -- SIPP reports compared with EHC reports -- we envision an analysis focused on missed events/circumstances in one or the other interviewing method ("missed" relative to each other and relative to administrative record data), errors (relative to records) in the reported timing of events and in income amounts received, and (where records are lacking) more or less reasonable distributions of spell transitions across calendar months. We would appreciate comments on this overall evaluation strategy and suggestions for additional analyses or other ways to improve it.

-- What additional evaluation methods might be employed to better understand the EHC interview process?

### **References**

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar. Potential improvements in the quality of retrospective reports in surveys. *Memory*, 6, 383-406.

Belli, R. F., Shay, W. L., & Stafford, F. P. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly*, 65, 45-74.

Callegaro, M. (2007). *Seam Effects Changes Due to Modifications in Question Wording and Data Collection Strategies, A Comparison of Conventional Questionnaire and Event History Calendar Seam Effects in the PSID*. Lincoln, NE: University of Nebraska

Fields, J. and Callegaro, M. (2007) Background and Planning for Incorporating an Event History Calendar into the Re-Engineered SIPP. Prepared for presentation at the Federal Committee for Statistical Methodology conference Nov. 2007.

Kominski, R. (1990). The SIPP Event History Calendar: Aiding respondents in the dating of longitudinal events. In *Proceedings of the Section of Survey Research Methods* (pp. 553-558). Washington D.C.: American Statistical Association.